ORIGIN OF THE DOUBLE HELIX AT THE CODON-ANTICODON INTERFACE

by

Brian K. Davis

Research Foundation of Southern California
8861 Villa La Jolla Dr. #13595, La Jolla CA 92037

(*published December 31, 2018*)

Principles applied to reconstruction of the evolution of replication and genetic code indicate that the transition from binary, RNA ladder-replicators, reliant on purine self-recognition, to the double helix, with complementary purine-pyrimidine base pairs, was driven by changes at the codon-anticodon interface. These changes enabled expansion of the code, increasing the array of encoded amino acid residues incorporated into proteins.

Key words: path-distance; codon time-order; path-invariant; replicator scaffold; translation

Anti-parallel pentose-phosphate strands in the scaffold of an RNA, or DNA, double-helix were recently interpreted to represent the conserved vestige of an earlier, ladder-replicator (Davis 2017, 2018a,b). Bi-directional H-bonds forming double-helix base pairs, exhibiting A $\rightarrow$ U/T $\rightarrow$ A, G $\rightarrow$ C $\rightarrow$ G complementarity[1] (Watson and Crick, 1953a,b), were noted, in addition, to conform with sequence propagation by self-recognition, as in A $\rightarrow$ A, G $\rightarrow$ G. A simple, direct mechanism of sequence propagation, consistent with ladder-replicator antiquity, resulted. The transition envisioned from propagation of a binary, linear base sequence to that of a quaternary, coiled sequence, with elevated complexity density (Fig. S1), conforms in a general sense with the anticipated direction of biological evolution.

Adenosine-rich codons assigned to diacid amino acids and their amides were found to formed the first code , which was identified after equating the time-order of base triplet assignments with amino acid synthesis path-distance (Figs. S2, S3). This finding furnished evidence that translation originated from the pre-code assembly of random sequence oligopeptides of these amino acids (Asp^1, Asn^2, Glu^1, Gln^2) on a poly(A) template (Fig. 1). Based on the path invariant principle derived during the former investigation (Davis, 2012, 2015), it became apparent, as noted, that ladder replicators, reliant on purine self-recognition, preceded Watson-Crick purine-pyrimidine complementarity and double helix. These findings linked transition from propagation of a binary RNA sequence, based on purine self-recognition in a ladder replicator, and quaternary purine/pyrimidine sequence replicators, to changes at the codon-anticodon interface, enabling expansion of the genetic code and its array of encoded amino acids.

[1] A, adenosine; G, guanosine; U/T, urIdine/thymIdine; C, cytodine; purines (A, G) predating pyrimidines (U(T), C).

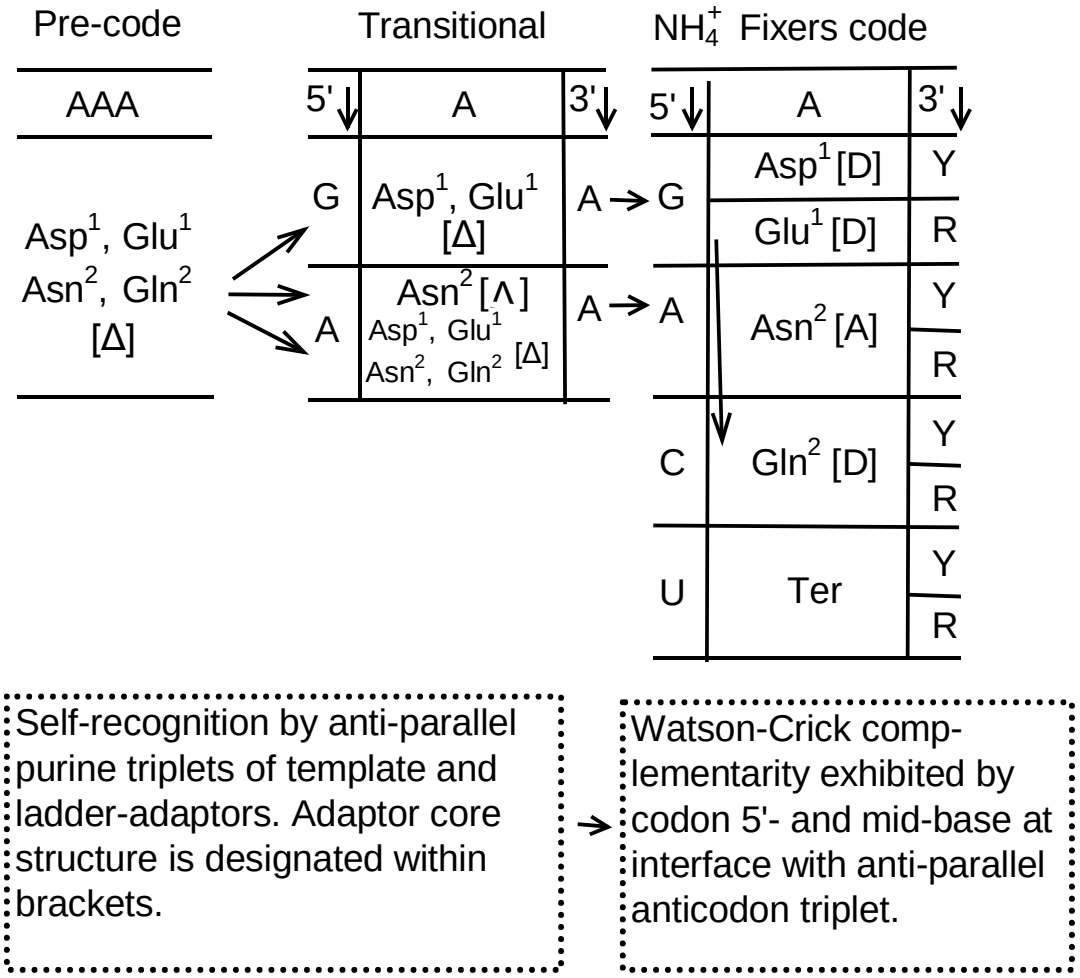
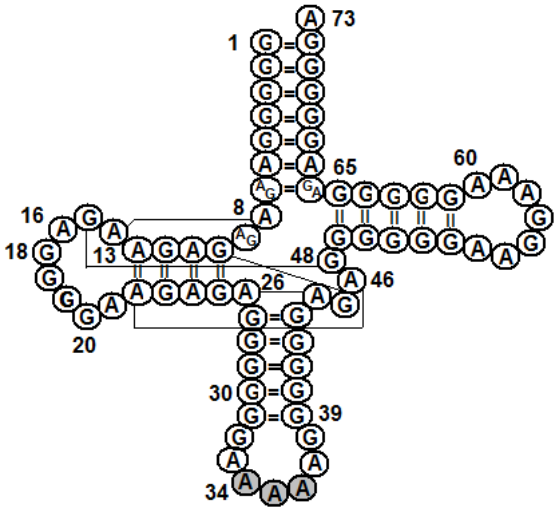
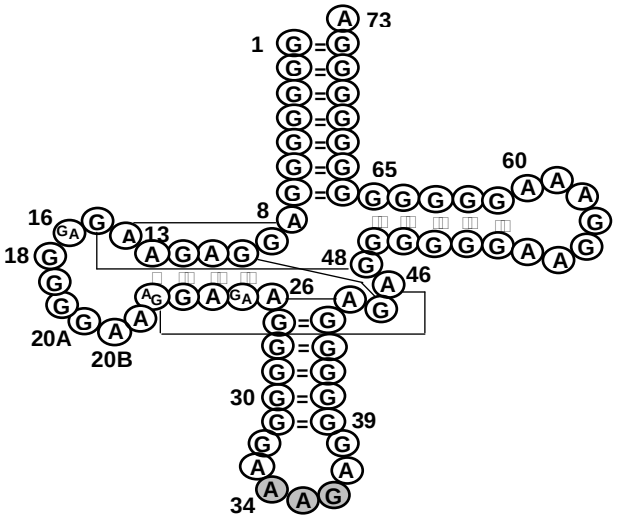


Fig. 1. Pre-code and transitional stages of translation, inferred from the NH_4^+ Fixers Code. Adaptor core structure is designated by Latin letters within brackets. In pyrimidine-deficient stages, the proto-adaptor core is represented by a Greek symbol. Asn^2 adaptor core, ʌ , differs from that of the core structure group, Δ, of the pre-code ladder adaptor, with which it competed for AAA codons.

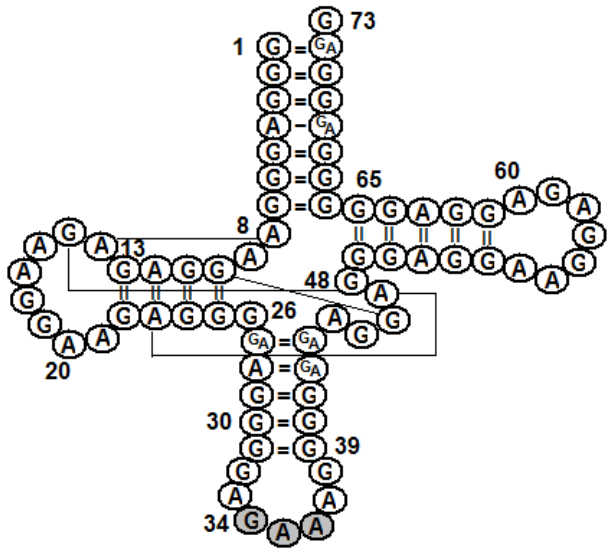
Clover-leaf structures of pre-code, diacid, and asparagine (Asn) adaptors in Fig.2 were constructed from consensus sequences for pre-divergence tRNA (Table S2). Retention of core structure groups throughout code formation, indicates the adaptors have not changed fundamentally over the intervening interval.



Pre-code Ladder-Adaptor [Δ]



Diacid Ladder-Adaptor [Δ]



Asn^2 Ladder Adaptor [Λ]

Fig. 2. Depicts clover-leaf structure of pre-code, diacid and asparagine ladder adaptors. These structures were deduced from consensus tRNA sequence given in Table S2.

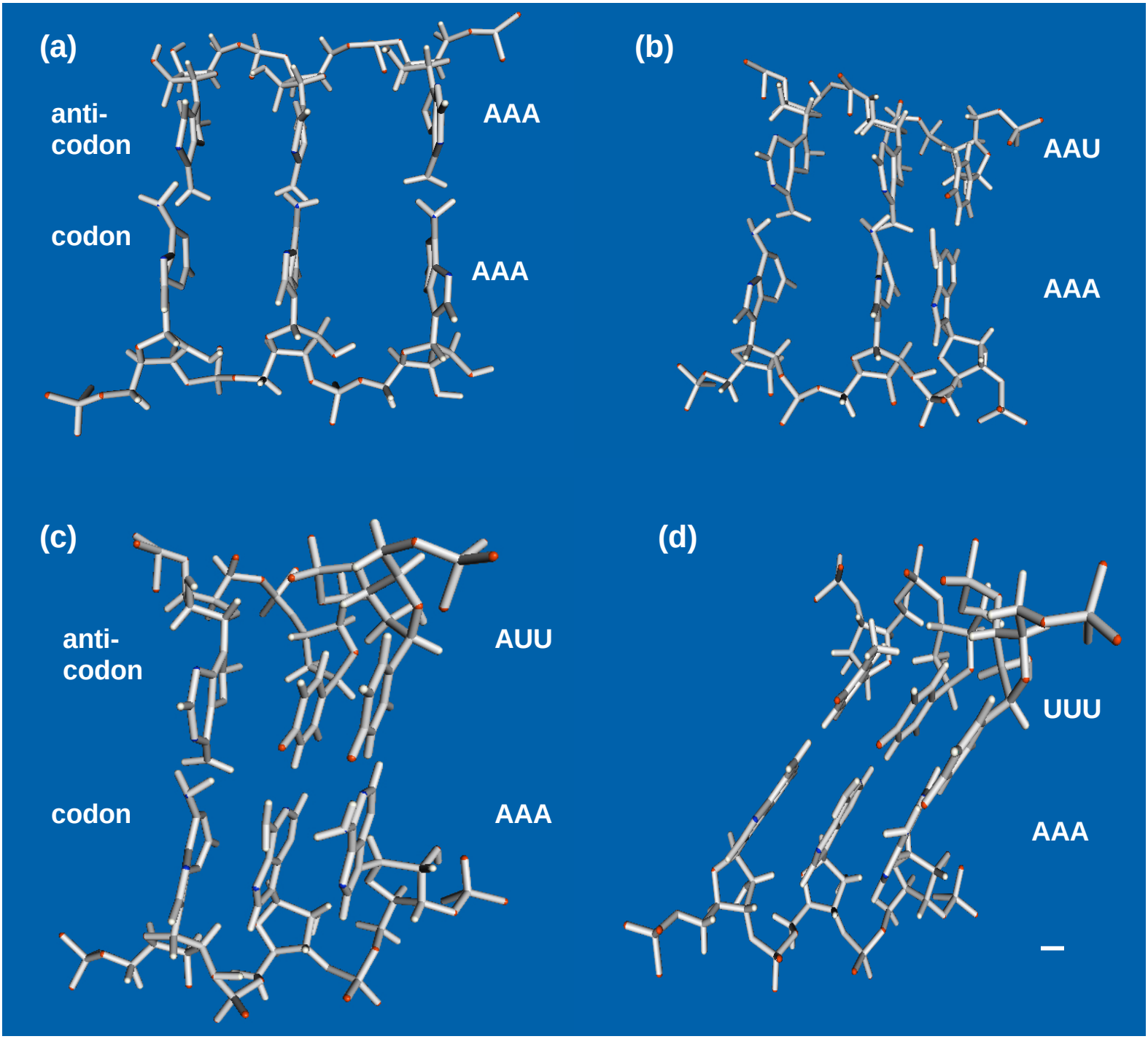


Fig. 3. (above) Shows transition transition from (a) ladder form of codon:anticodon pair AAA:AAA to transitional pairs (b) and (c) to type A RNA double helix in (d) with complete displacement of the initial AAA anticodon by the complementary pyrimidine, U.

The ladder to double helix transition depicted in Fig. 3 with the replacement of A by its complementary pyrimidine in a stepwise manner from adaptor N34 to N36, accompanied increasing coding specificity. Coevolution of the proof-reading purines in the decoding center could be anticipated.

Supplement

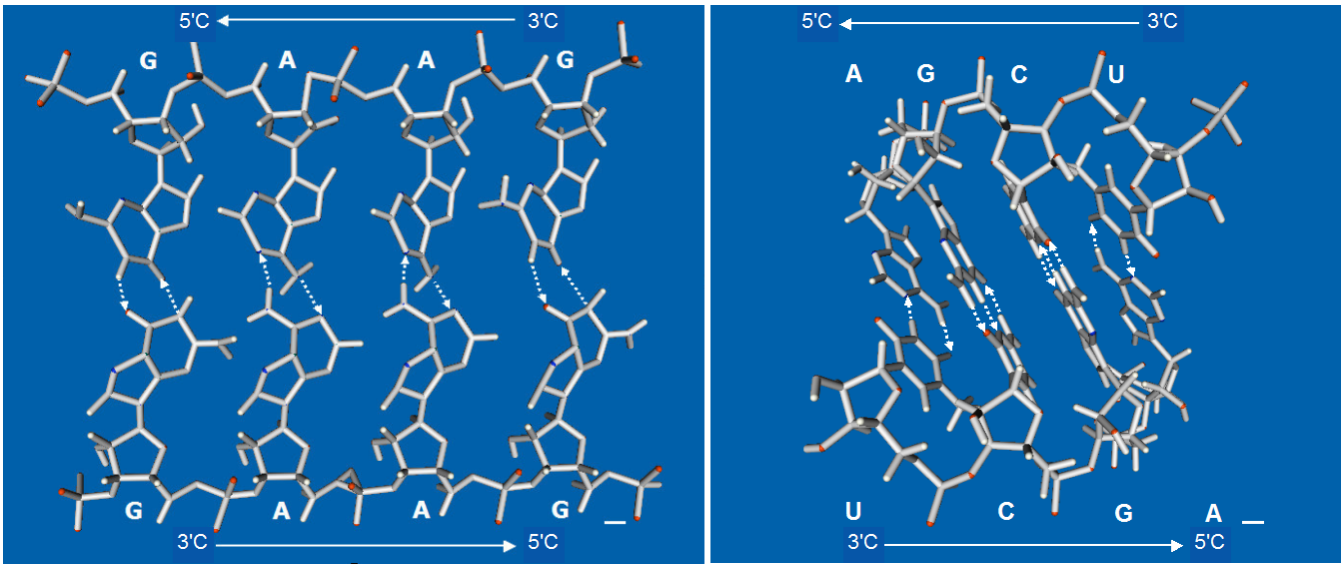


Fig. S1. Comparison between RNA tetramers in a purine-ladder (a) and type A double-helix (b). Purine self-recognition pairs characterize the former, and Watson-Crick purine-pyrimidine complementarity governs pair formation in the latter. Arrows between designated bases indicate base pair H-bonds are bi-directional. Upper and lower arrows refer to the direction of the anti-parallel poly(ribose-phosphate) scaffold. Bar length, 1 Å.

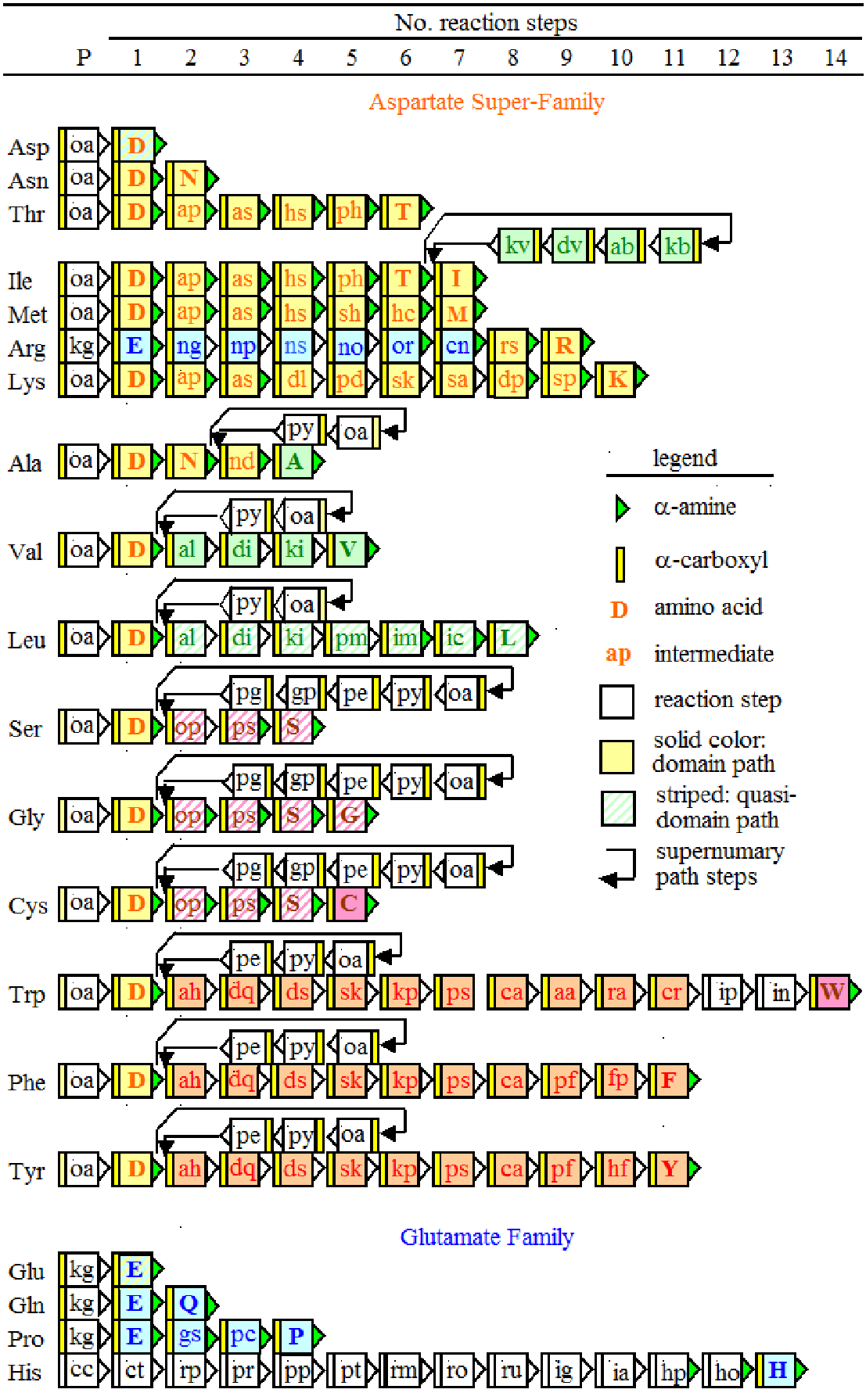

No. reaction steps
P 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Aspartate Super-Family
Asp oa D
Asn oa D N
Thr oa D ap as hs ph T
kv dv ab kb
Ile oa D ap as hs ph T I
Met oa D ap as hs sh hc M
Arg kg E ng np ns no or cn rs R
Lys oa D ap as dl pd sk sa dp sp K
Ala oa D N nd A py oa
Val oa D al di ki V py oa
Leu oa D al di ki pm im ic L py oa
Ser oa D op ps S pg gp pe py oa
Gly oa D op ps S G pg gp pe py oa
Cys oa D op ps S C pg gp pe py oa
Trp oa D ah dq ds sk kp ps ca aa ra cr ip in W pe py oa
Phe oa D ah dq ds sk kp ps ca pf fp F pe py oa
Tyr oa D ah dq ds sk kp ps ca pf hf Y pe py oa
legend
α-amine
α-carboxyl
D amino acid
ap intermediate
reaction step
solid color: domain path
striped: quasi-domain path
supernumary path steps
Glutamate Family
Glu kg E
Gln kg E Q
Pro kg E gs pc P
His cc ct rp pr pp pt rm ro ru ig ia hp ho H

Fig. S2. Amino acid path-distances in reconstructed tRNA-dependent synthesis pathways. Number of reaction steps appear in the overbar. Oxaloacetate (oa) was precursor to fifteen amino acids, forming an enlarged, pre-enzyme aspartate super-family. Ketoglutarate (kg) produced four glutamate family amino acids and ribose-5-phosphate (rp) produced one. cc, refers to citrate cycle metabolite; and ct, to central trunk. An intermediate lacking a tRNA cofactor attachment site is represented by a white background. Three-letter amino acid abbreviations occur in the left-hand column; upper-case, single-letter amino acid abbreviations occur within pathways. Lower-case, double-letter abbreviations denote non-amino-acid intermediates (Michal, 1992)**:** py, pyruvate, pe, phosphoenolpyruvate, gp, 2-phosphoglycerate, pg, 3-phosphoglycerate. **Thr** - ap, aspartyl-phosphate; as, aspartate-β-semialdehyde; hs, homoserine; ph, o-phospho-homoserine. **Ile** - kb, α-keto-butyrate; ab, α-aceto-α-hydroxy-butyrate; dv, α,β-dihydroxy-isovalerate; kv, α-keto-isovalerate. **Met** – sh, o-succinyl-homoserine; hc, homocysteine. **Arg** - ng, N-acetyl-glutamate; np, N-acetyl-glutamate-phosphate; ns, N-acetyl-glutamate-γ -semialdehyde; no, N-acetyl-ornithine; or, ornithine; cn, citrulline; rs, arginine-succinate. **Lys** - dl, α,β-dihydropicolineate; pd, Δ^1-piperdiene-2,6-dicarboxylate; sk, N-succinyl-ε-keto- α-amino-pimelate; sa, N-succinyl- α,ε-diamino-pimelate; dp, α,ε-diamino-L-pimelate; sp, meso- αε-diamino-pimelate. **Ala** - nd, Glu amine-donor; Asn-like cofactor/adaptor. **Val** - al, α-aceto-lactate; dl, α,β-dihydroxy-isovalerate;kl, α-keto-isovalerate. **Leu** - pm, α-isopropyl-malate; im, β-isopropyl-malate; ic, α-keto-isocaproate. **Ser** – op, phospho-hydroxypyruvate; ps, phospho-serine. **Trp** - ah, β-deoxy-arabino-heptulosonate-7-phosohate; dq, 5-dehydroquinate; ds, 5-dehydro-shikimate; sk, shikimate; kp, shikimate-5-phisohate; ps, 3-enolpyruvyl-shikimate-5-phosphate; ca, chorismate; aa, anthranilate; ra, N-phospho-ribosyl-anthranilate; cr, 1-(o-carboxyphenylamino)-1'-deoxyribulose-5-phosphate; ip, indole-3-glycerol-phosphate; in, indole. **Phe** - pf, prephenate; fp, phenyl-pyruvate. **Tyr** - hf, p-hydroxy-phenyl-pyruvate. **Pro** - gs, glutamate-γ-semialdehyde; pc, Δ1-pyrroline-5'-carboxylate. **His** - pp, phosphatidyl-ribosyl-pyrophosphate; pt, phospho-ribosyl-adenosine-triphosphate; rm, phospho-ribosyl-adenosine-monophosphate; ro, phospho-ribosyl-formimino-amino-imidazole-carboxamide-ribose-phosphate; ru, phospho-ribulosyl-formimino-amino-imidazole-carboxamide-ribose-phosphate; ig, erythro-imidazole-gylcerol-phosphate; ia, imidazole-acetol-phosphate; hp, histidinol-phosphate; ho, histidinol.

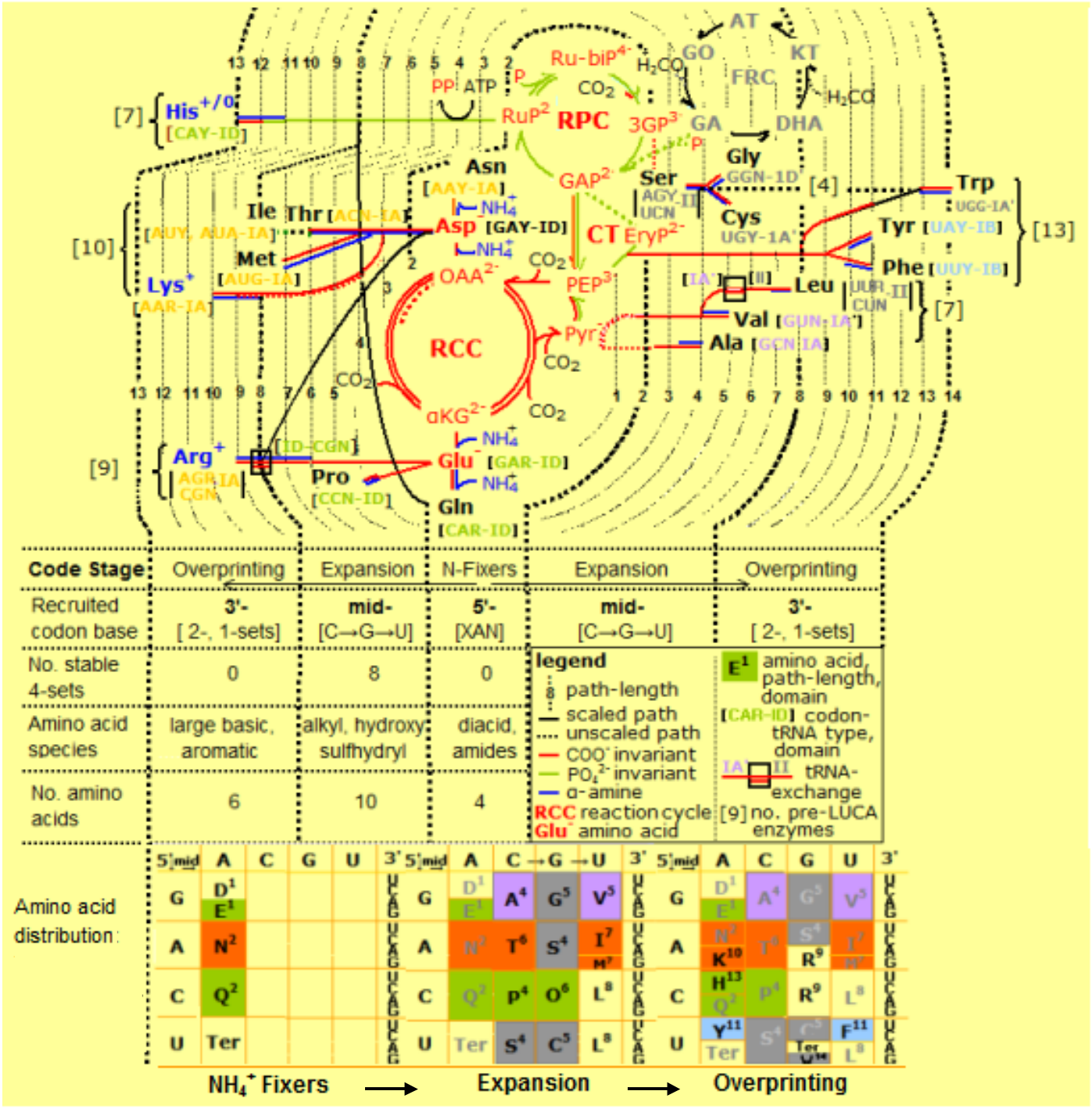


Fig. S3. Depicts relationship between extension of tRNA-dependent amino acid synthesis pathways, from central metabolism, with inferred time-order of codon assignments during formation of the genetic code. Adapted from Davis (2018).

Table S1. Known and new structural features encompassed by the path-distance model of genetic code evolution. A comprehensive list of code features, with interpretations, appears in references under this table.

GENETIC CODE FEATURES UNIFIED BY PATH-DISTANCE MODEL

Known	New
1. Woese, 1965: NAN, NUN aa hydropathy clusters.	1. Code comprises six domains of contiguous codons read by related pre-LUCA tRNA for same-family aa.
2. Bretscher et al. 1965: nonsense codon inhibition.	2. Amino acid synthesis intermediates retain an invariant α-carboxyl linked to early tRNA-cofactor attachment.
3. Nirenberg et al. 1966: aa biosynthetic clusters.	3. Path-distances reveal codon bases were assigned to distinct kinds of aa in 5'→mid→3' order.
4. Dunnill, 1966: codon 4-sets have 5'/mid-G, C.	4. Compact XAN codon set (X, coding site) first encoded four N-fixer aa (1-2 step paths) and a stop signal.
5. Crick, 1966: universality of standard code.	5. First code places origin of proteins at two RCC N-fixer sites, yielding diacids Asp^1, Glu^1 and amides Asn^2, Gln^2.
6. Wilcox, Nirenberg, 1968: tRNA an aa cofactor.	6. Source duality is amplified in diacid function (aa source v. N donor), phylognetic depth, and synthetase class.
7. Rodwell, 1969: Ile^7 path has four Val^5 steps.	7. Pre-LUCA tRNA identities indicate Asp^1 was initially precursor to 15 aa, and Glu^1 to only 3 aa.
8. Dillon, 1973: 4-sets predated 2- and 1-sets.	8. Asn^2 and Gln^2 attachment to tRNA blocked lactam formation by these early, labile aa.
9. Dillon, 1973, Wong, 1975: aa coevolved with tRNA.	9. Mid-base was assigned in an (A)→C→G→U order to ten increasingly hydrophobic aa of ~4, 5, and 7 path-steps.
10. Dillon, 1973, Wächtershauser, 1992: aa synthesis paths formed by reductive organo-synthesis.	10. Eight stable code-boxes (4-sets) were assigned (GCN to Orn^6) during expansion from the N-fixer's code.
11. Perlwitz et al. 1988: mid-base most coding capacity.	11. 3'-Base encoded six basic/aromatic aa of 9-14 path-steps, by overprinting six error-prone boxes.
12. Taylor, Coates, 1989: sibling aa share codon 5'-base.	12. tRNA-cofactor exchange led to anomalous assignment of UUR, CUN to Leu^8, and AGR, CGN to Arg^9.
13. Taylor, Coates, 1989: smallest aa assigned 4-sets.	
14. Garrett, Grisham, 1999: na-like aa have long paths.	
15. Lim, Curran, 2001: Y:Y wobble split eight c4-sets.	
16. Brooks et al. 2002: ancient proteins have early aa	
17. Brooks et al. 2002: early aa in ancient proteins.	
18. Brooks, Fresco, 2003: GNN code for early aa.	
19. Biro et al. 2003: codon R, Y mid-base aa clusters.	
20. Norgaard et al. 2009: reconstruction of Pro-Fd-5.	
21. Rodin et al., 2009. tRNA N2:N71 complementarity.	
22. Williams et al. 2009: Synhetase duality.	

Davis, BK 2007. Making sense of the genetic code with the path-distance model. In, *Leading-Edge Messenger RNA Research Communications* Ed., MH Ostrovskiy. New York: Nova Science, Chp. 1, pp. 1-32.
Davis, BK 2013. Making sense of the genetic code with the path-distance model based on tRNA-dependent pathways. https://archive.org/details/MakingSenseOfGeneticCode DOI: 10.13140./RG.2.2.17217.86888

Table S2. LUCA tRNA sequence source: Davis B. K. 2008. Imprint of early tRNA diversification on the genetic code: Domains of contiguous codons read by related adaptors for sibling amino acids. In *Messenger RNA Research Perspectives* Ed. T. Takeyama. New York: Nova Biomedical Books Chp. 1, pp. 35-79.

Amino acid specificity	Adaptor	Coding triplet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17A	18	19	20	20A	20E	21	22	23	24	25	26	27	28	29	30
Asp[1]	tRNA-D	3'-CUG	G	C	C	C	U	G	G	U	A	G	U	G	U	A	G	U	CU		G	G	U	C	N	A	G	N	A	U	A	C	G	N	G
Glu[1]	tRNA-D	3'-CUU	G	C	C	C	C	N	N	U	G	G	U	G	U	A	G	C			G	G	C	CU		A	U	C	A	C	A	C	C	G	C
		3'-CUC	G	C	C	C	C	C	G	U	G	G	U	G	U	A	G	CU			G	G	C	C	AU	A	GU	GC	A	CU	A	CU	C	G	C
Gln[2]	tRNA-D	3'-GUU	N	G	N	C	C	U	A	U	A	G	U	G	U	A	G	CU			G	G	U	CU		A	U	C	A	C	N	N	C	G	G
		3'-GUC	N	G	U	C	C	U	A	U	N	G	U	G	U	A	G	U			G	G	C	CU		A	U	C	A	C	G	AU	C	G	G
	Consensus tRNA:		G	C	C	C	C	U	AG	U	AG	G	U	G	U	A	G	U			G	G	C	C		A	U	C	A	C	A	C	C	G	G
Asp[1], Asn[2] Glu[1], Gln[2]	Ladder-Δ	AAA	G	G	G	G	G	A	AG	A	AG	G	A	G	A	A	G	A			G	G	G	G		A	A	G	A	G	A	G	G	G	G
Asp[1], Glu[2]	Cons. tRNA	3'-CU$^{G}_{Y}$	G	C	C	C	C	$^{G}_{C}$	G	U	G	G	U	G	U	A	G	$^{U}_{C}$			G	G	C	C	$^{A}_{U}$	A	$^{U}_{G}$	C	A	$^{U}_{C}$	A	C	C	G	C
Asp[1], Glu[2]	Ladder-Δ	GAA	G	G	G	G	G	G	G	A	G	G	A	G	A	A	G	$^{A}_{G}$			G	G	G	G	A	A	$^{G}_{A}$	G	A	$^{G}_{A}$	A	G	G	G	G
Asn[2]	tRNA-A	3'-UUG	G	C	C	U	C	C	G	U	A	G	C	U	C	A	G	U	U		G	G	U			A	G	A	G	C	G	N	U	C	G
Asn[2]	Ladder-Λ	AAA	G	G	G	A	G	G	G	A	A	G	G	A	G	A	G	A	A		G	G	A			A	G	A	G	G	G	$^{G}_{A}$	A	G	G

Coding triplet	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
	LUCA NH_4^+-Fixers tRNA-D																																										
3'-CUG	C	C	U	G	U	C	A	C	G	C	N	C	G	A	G	A		C	N	C	G	G	G	U	U	C	A	A	N	U	C	C	C	G	G	C	C	G	G	G	G	C	G
3'-CUU	C	C	U	U	U	C	A	C	G	G	C	G	G	N	G	A		C	C	G	G	G	G	U	U	C	G	A	A	U	C	C	C	C	G	N	N	G	G	G	G	C	A
3'-CUC	C	C	U	G	U	C	A	C	G	G	C	G	G	U	G	AG		C	CG	C	G	G	G	U	U	C	G	A	U	U	C	C	C	G	G	C	G	G	G	G	G	N	A
3'-GUU	A	U	U	U	U	G	A	A	N	C	C	G	N	C	A	A		C	C	U	A	G	G	U	U	C	G	A	A	U	C	C	N	G	G	U	A	G	G	N	C	N	A
3'-GUC	A	U	U	C	U	G	G	N	U	C	C	G	AU	U	A	A		C	C	C	A	G	G	U	U	C	G	A	G	U	C	C	U	G	G	U	A	G	G	A	C	U	G
RNA:	C	C	U	G	U	C	A	C	G	C	C	G	G	U	G	A		C	C	C	G	G	G	U	U	C	G	A	A	U	C	C	C	G	G	CU	A	G	G	G	G	C	A
	Pre-code Adaptor																																										
AAA	G	G	A	A	A	A	A	G	G	G	G	G	G	A	G	A		G	G	G	G	G	G	A	A	G	G	A	A	A	G	G	G	G	G	AG	A	G	G	G	G	G	A
	Proto-NH_4-Fixers Adaptors																																										
3'-CU^{G}_{Y}	C	C	U	G	U	C	A	C	G	C	G	G	G	$^{U}_{A}$	G	A		C	C	C	G	G	G	U	U	C	G	A	$^{U}_{A}$	U	C	C	C	G	G	C	$^{G}_{C}$	G	G	G	G	C	A
GAA	G	G	A	G	A	G	A	G	G	G	G	G	G	A	G	A		G	G	G	G	G	G	A	A	G	G	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	A
3'-UUG	G	C	U	G	U	U	A	A	C	C	G	N	N	A	G	G	U	C	G	C	A	G	G	U	U	C	G	A	G	U	C	C	U	G	C	C	G	G	N	G	G	AC	G
AAA	G	G	A	G	A	A	A	A	A	G	G	G	$^{G}_{A}$	$^{G}_{A}$	A	G	G	A	G	G	G	A	G	G	A	A	G	G	A	G	A	G	G	A	G	G	G	G	G	$^{G}_{A}$	G	G	$^{G}_{A}$	G

Table S2 (continued)

References

Davis BK 2012. Replicative-form of poly(triose-phosphate).
archive.org/details/Replicative-formOfPolytriose-phosphate

Davis BK 2015. Path invariants and the evolution of replication.
archive.org/details/EvolnOfRepln2015

Davis BK 2017. Hetero-Polypentotide as a Pre-RNA Replicator: Origin of Non-Local Order
archive.org/details/Pre-rnaReplicators DOI: 1013140/RG.2.2.10499.40487

Davis BK 2018a.Deep Structure of the Genetic Code and the Origin of Replication: Path Invariants as Pre-LUCA Attachment Sites. archive.org/details/PreRNAReplicators DOI: 1013140/RG.2.2.20657.68967

Davis BK 2018b. Origin of Watson Crick Complementarity.
archive.org/details/OriginOfWatsonCrickComplementarity DOI: 1013140/RG.2.2.25292.72322